Authors: David T. Hansen, Chris Curlis, and Barbara D. Simpson

Geodatabase for Evaluating Imagery and Other Data

Presented by David T. Hansen at the ESRI User Conference, 2006, San Diego California, August 8, 2006

Abstract

A common tool in the evaluation and use of remote sensing data is the use of polygons for the comparison of spectral characteristics from satellite imagery. A geodatabase provides a useful structure and topology tools for evaluating polygons derived from spectral imagery with other data such as census tracts, administrative boundaries, water areas, or field borders. This office is using the geodatabase structure and associated tools for assessing spectral variations in the dynamic environment of the Central Valley of California. Geodatabases with ArcMap provide the basis for evaluating image information captured with Definiens software eCognition. The geodatabase provides the flexibility to rapidly address a variety of issues. This includes identifying land cover changes near facilities or on Reclamation administered lands, the status of lands involved in fallowing programs, and base line information for water conservation programs.

Introduction

The U.S. Bureau of Reclamation is evaluating land cover change in the Central Valley, California. There are several tasks associated with land cover mapping in the Central Valley. These include generation of a dataset of polygons representing land cover features, adjustment and integration of these polygons with other land cover information. The base polygon dataset for the first round of land cover change for the time period 1993 to 2000 was generated from image segmentation of Landsat 5 TM imagery with a resolution of 30 meters. This base polygon set was not modified but used to capture additional land cover information with raster processing. This initial change project is described in several publications including development of the land cover legend and capture of base land cover information by Hansen et. al. (2001), and the identification of change between 1993 and 2000 by Curlis, et. al. (2003). This initial round of change detection focused on change affecting Reclamation water service areas. It assisted in identifying areas for habitat acquisition and enhancement for wildlife species of concern in the Central Valley.

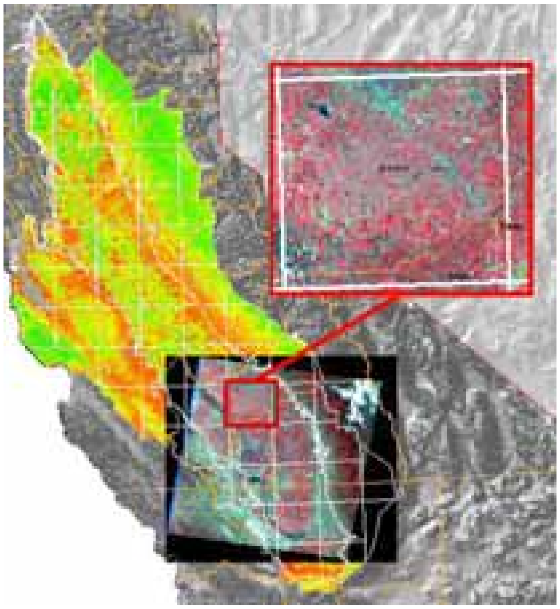
The effort for 2005 is focused on change near Reclamation facilities and areas served by Reclamation as well as identifying change in major land cover types throughout the valley. Of particular interest are lands involved in fallowing programs and land use changes that may affect water conservation plans. For the 2005, pan sharpened Landsat 7 TM imagery is being used. Image segmentation on the pan sharpened imagery results in a finer set of polygons that more closely represent features observed in high altitude aerial photography. This is particularly evident for man made features such as agricultural fields, roads, canal systems, large structures, and other developed areas. Such features were observed in the 2000 base polygon set but no adjustments were made in the position or configuration of the spectral polygons. The 2005 base polygon set is being adjusted to more closely match these features. Adjustment to the base polygon dataset includes the incorporation of more detailed sets of polygons generated from source data or imagery.

New tools including ArcMap 9+, geodatabase models, geoprocessing tools and Definiens eCognition software are assisting in this effort. This paper describes the development of a geodatabase model for maintaining and editing the polygons generated for the 2005 update along with the associated feature classes, raster data, and related tables. The ArcGIS geoprocessing framework with geoprocessing tools will be used for maintaining and updating the datasets. On the release of ArcGIS 9.2, it is expected that the initial model will migrate from a personal geodatabase into a file geodatabase structure.

Overview of the Process for 2005 Land Cover Analysis, Central Valley, California

The Central Valley covers approximately 4,795,000 hectares (11,850,000 acres) of California. The overall program area covers an additional 5,000,000 hectares (13,000,000 acres) of the Sierra Nevada and California coast ranges to address areas near or served by Reclamation facilities. To cover such a large extent, the area is divided into 58 processing areas of about 180,000 hectares (453,000 acres) in size.

The processing for the 2005 land cover product is split into three general levels which have different spatial extents. These are processing at the full extent of the study area, processing of separate Landsat 7 TM scenes, and processing within the individual processing areas. Figure 1 shows the relative footprints for these spatial extents. Most ancillary data in raster or vector format can be processed for the full study area. Processing of Landsat scenes for the base spectral and change polygon datasets can currently only effectively be done at the scene or partial scene extent. Some image analysis and classification requires smaller subsets of data at a processing area level. A processing area is high lighted in the figure. Processing areas are also used to subset datasets for review and editing.



Change between 2000 and 2005 is developed from Landsat 7 ETM+ SLC off – SLC off Gap filled USGS data products. Gaps in the scene are filled with earlier or later Landsat passes. Image differencing techniques are used to produce a change product of spectral differences between 2000 and 2005 imagery. This differencing is done in Definiens eCognition software based on the segmented image created from a multi date composite image. Thresholds are set individually to create three spectral change images. Each TM band difference component identifies somewhat different types of spectral change. The three band images are combined to identify the full range of spectral change. Shape files are generated from eCognition on a full scene basis representing the full set of spectral polygons and the subset of change polygons. Multiple Landsat scenes are required to cover the area of interest.

Ancillary data with image processing are being used to isolate and identify areas of true change from change related to differences in cultural practices or climate conditions between the image dates. This reduces the number of spectral change polygons in the change polygon set for review and editing. Much of this spectral change is due to agricultural practices. MODIS NDVI 250 meter data products and other crop information are used to mask and eliminate change polygons that are due solely to active agriculture (Hansen, et. al., 2006). Other data such as current crop mapping, hydrography and more detailed image segmentation and classification are used to update the base polygon dataset.

Ancillary data representing roads and canal systems are also used to associate additional information on land cover or land use with the base polygons. The current set of ancillary data for 2005 is listed in the appendix. The land cover project for 2005 is an open process which can incorporate more detailed analysis of land cover categories for particular areas of interest.

Geodatabase Model for Land Cover Analysis, Central Valley, California

The geodatabase is intended to manage the data and control the processing for the land cover product for 2005 and change identified from 2000. It contains portions of spatial features and attributes from source datasets that are useful for the analysis. The geodatabase structure assists in managing and controlling several tasks.

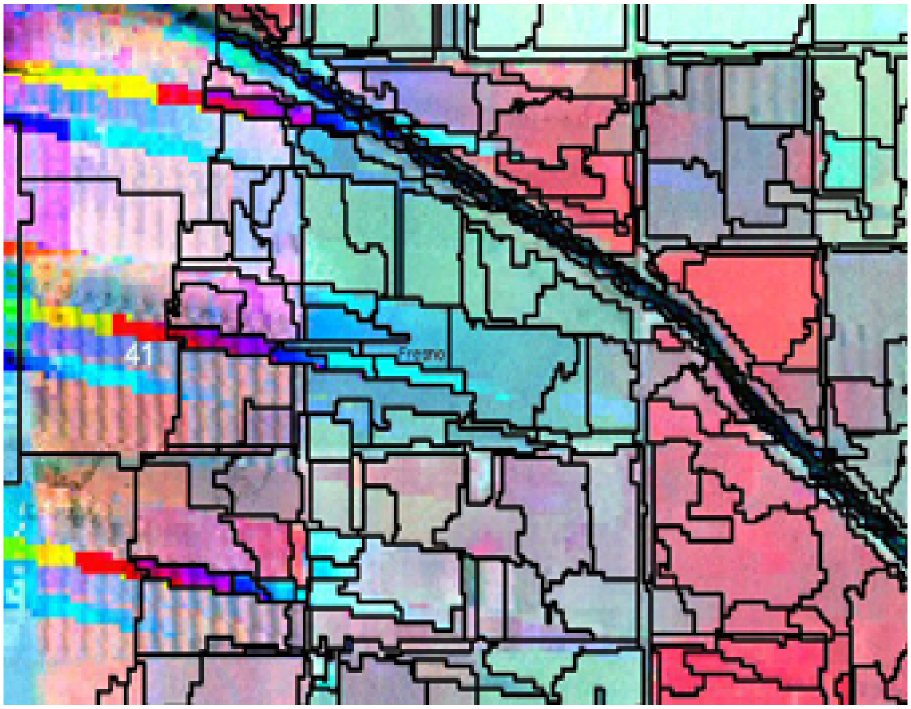
Basic Processing Tasks

The major tasks for the 2005 land cover in a geodatabase framework are:

- Maintaining a common spatial reference system.
- Merging polygons from adjacent and overlapping datasets.
- Capture of other land cover information.
- Updating geometry and attributes of the base polygon datasets.
- Adjusting the attributes and geometry of polygon datasets.
- Associating base polygons with other sources of land cover or land use.
- Managing linkage between feature datasets to attributes and to source data.
- Incorporating common tasks in geoprocessing scripts and tools

Maintaining a common spatial reference system - The geodatabase enforces a common spatial reference for all spatial data. This is particularly important for this project where data is continually being moved between different software packages. A primary package that is being used for image segmentation and classification, Definiens eCognition, does not read or write projection information. It will read in thematic data as shape files and will export the results of object classifications as shape files, but without spatial reference information. Loading these shapes into the geodatabase ensures that a proper spatial reference system is assigned. In addition, raster datasets are maintained in the same coordinate space for raster analysis and zonal statistics with the base polygon datasets. This is currently independent of the personal geodatabase in ArcGIS 9.1. With the release of ArcGIS 9.2, we hope to have raster data stored with vector data with a common spatial reference system.

Merging polygons from adjacent and overlapping datasets - The basic process flow for the 2005 land cover mapping project begins with generation of spectral polygons and related change polygons from the Landsat 7 TM pan sharpened imagery. These are developed at a full or partial TM scene extent dependent on processing limitations. As a part of the process, polygons from adjacent scenes are merged with adjacent scenes. Scene overlap ensures that spectral polygons will overlay those from adjacent scenes. This overlap can be beneficial due to scan line drop and fill issues for the Landsat 7 data. Adjacent scene polygons may more appropriately represent the area. Figure 2 shows a portion of polygons generated in eCognition from a Landsat 7 TM scene for June 2005.



The geodatabase structure assists in the review and merging of overlapping polygons with topology rules. Adjacent scenes are managed as separate subtypes representing different source data with topology rules on overlapping polygons and default values dependent on the source. Validation based on these rules should assist in merging the adjacent polygon datasets.

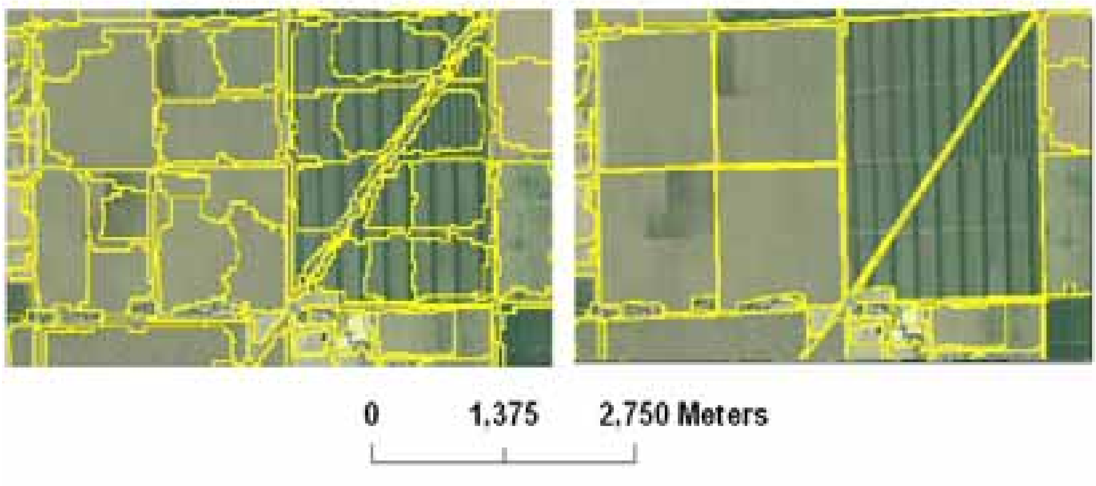
Capture of other land cover information - While ancillary datasets are available for the full study area, only portions of the data may be used in processing at the TM scene or processing area level. Generally only subsets of both the spatial features and attributes are needed from these different sources. Frequently source data have gaps in the spatial extent. Sometimes the source data do not represent the same time period. Some data are expected to have updates during the development of the 2005 land cover dataset. The geodatabase structure permits loading of the selected set of spatial features and attributes from the different sources. As updates occur, these can be incorporated into the geodatabase without having to alter the structure. Separate tables are maintained in the geodatabase with source attribute information for linkage to the vector feature datasets with key fields.

Currently, the following information is associated with the spectral polygons:

- Land cover classification from 2000.
- Active agricultural areas for 2005 based on MODIS NDVI 250 meter data products.
- Open water based on National Hydrography Dataset (NHD) and other detailed hydrography feature classes.
- Known natural areas for 2005 such as wildlife areas or preserves.

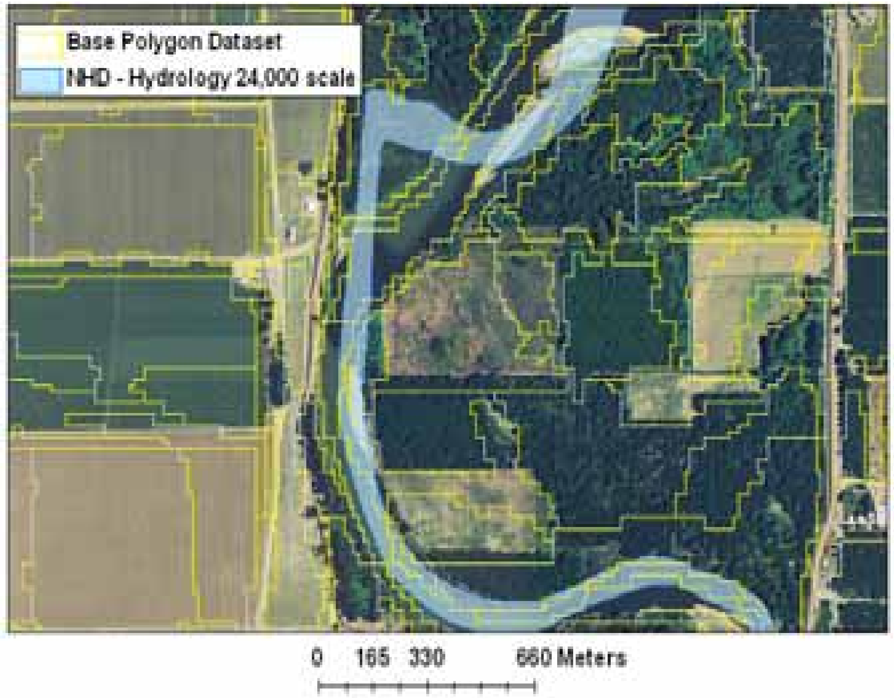
The capture or assignment of information from the source land cover data for the spectral polygons is done via geoprocessing tools. Much of this capture is done in raster format. Information from the separate sources is captured with zonal statistics for the base polygons as table datasets.

Updating geometry and attributes of base polygon datasets - In some cases, the set of polygons from a source represents features better than the existing polygon dataset. An example is crop information from USDA, Farm Services Agency (FSA). In this case, the spectral polygons are updated with the set of polygons from the crop data. Figure 3 show an area with the original spectral polygon dataset on the left and the resulting update representing crop information from FSA on the right. Updating the base polygon dataset with the polygons from the FSA crop data improves the representation of fields and simplifies the set of polygons.

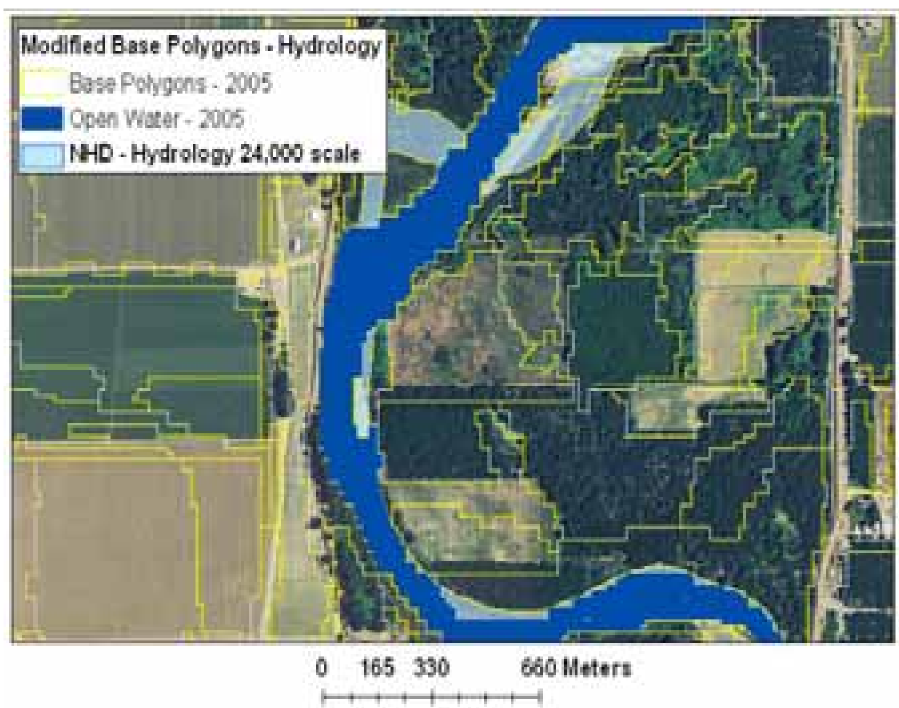


The geodatabase assists in controlling this update process by the use of scripting with geoprocessing tools. For the update process, the input polygon datasets have a simple set of attributes which include the source and a unique identifier for the polygon. Other attribute information are maintained as independent tables.

Adjusting attributes and geometry of polygon datasets - The processing area level represents a subset of polygons for review and editing. Typically, this review is with one meter 2005 NAIP color ortho imagery. Besides identifying actual change between 2000 and 2005, the land cover classification is reviewed. Review with the one meter imagery will indicate where adjustments in the geometry of the base polygon dataset may be desired. An example of this is shown in Figure 4. This figure shows an area along the Sacramento River where the channel has changed from the date that the channel was originally mapped as represented in NHD and detailed hydrography. Integration, with identity, of the spectral polygons with detailed hydrography permits a better representation of this open water feature.

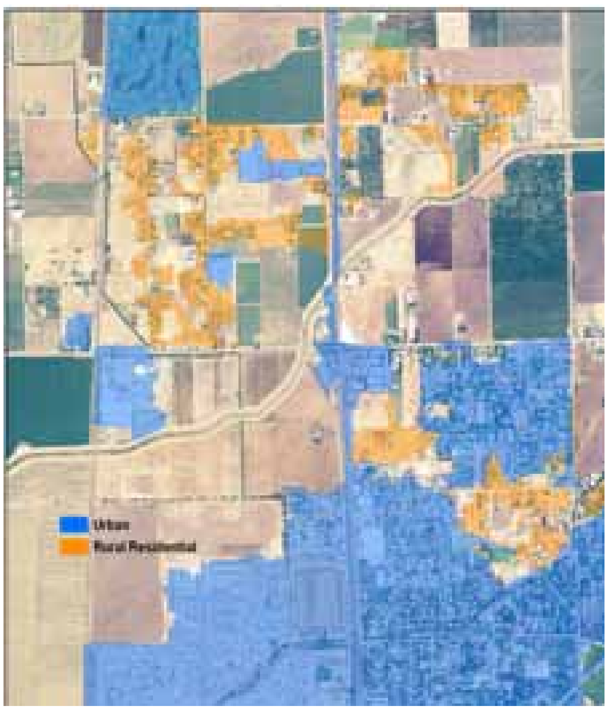


For this area decisions are made on which source represents better geometry for the combined set of polygons. For the area shown in Figure 4, a portion of the Sacramento River channel from the source hydrography dataset is integrated with the set of base polygons. Where the channel has actually shifted since the date of the NHD data, attributes are edited to carry the classification of open water. These adjustments are shown in Figure 5.

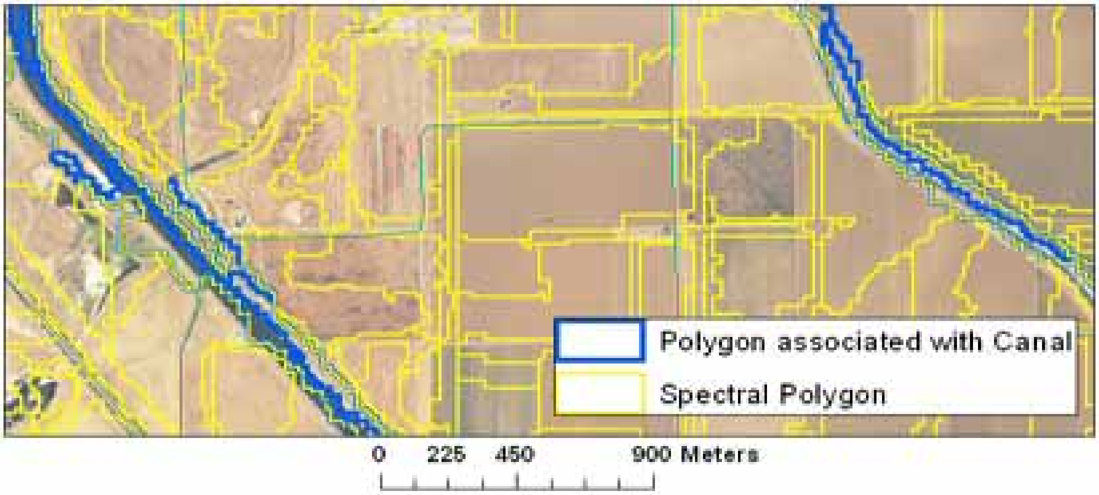


The attributes carrying the assignment of land cover for the polygons are maintained in a separate table. The geodatabase structure simplifies the task of editing the land cover assignment on this related table.

Associating base polygons with other sources of land cover or land use - Typically, additional multi-resolution segmentation and classification of land cover categories are done in eCognition at the processing area level. This includes identification of the urban foot print and fragmentation due to rural residential development. Rural residential developments were recognized but were difficult to handle in the land cover classification system for 2000. For these land use categories, the TM scene is combined with Census block information in eCognition to generate a subset of polygons of urban and rural residential that contain a mix of land cover categories. This is incorporated into the base polygon dataset (Simpson et. al., 2006). An example of an area categorized as urban and rural residential is shown in Figure 6.



The pan sharpened 15 meter data results in a large number of linear spectral polygons that follow man made features such as roads and canals. Like rural residential developments, these features consist of a complex set of land cover categories that include the structures that produce these patterns. These have been a problem in categorizing land cover. Figure 7 shows an area of the Central Valley with canals with the eCognition set of generated base polygons. Spatial selection of the base polygons with the road and canal networks can associate a broader land use category with these polygons to assist in review. Similar to rural residential this information can be maintained in separate tables in the geodatabase for future classification decisions.



Managing linkage between feature datasets to attributes to source datasets - The base polygon datasets have a simple attribute structure describing the source of the polygon and a unique identifier. These fields are used to link to independent tables which carry associated information such as land cover classification and land use categorization. Separate tables also provide the information needed to identify and link to source data. The geodatabase framework keeps this data together and identifies the relationships between the feature datasets and the table datasets.

Incorporating common tasks in geoprocessing scripts and tools - The geodatabase framework provides a basis for developing geoprocessing tools to consistently process the data subsets. Scripting with these tools permits the documentation of workflow and sequence of processes completed for the Landsat scenes and individual processing areas.

Conceptual Model for the Management Geodatabase

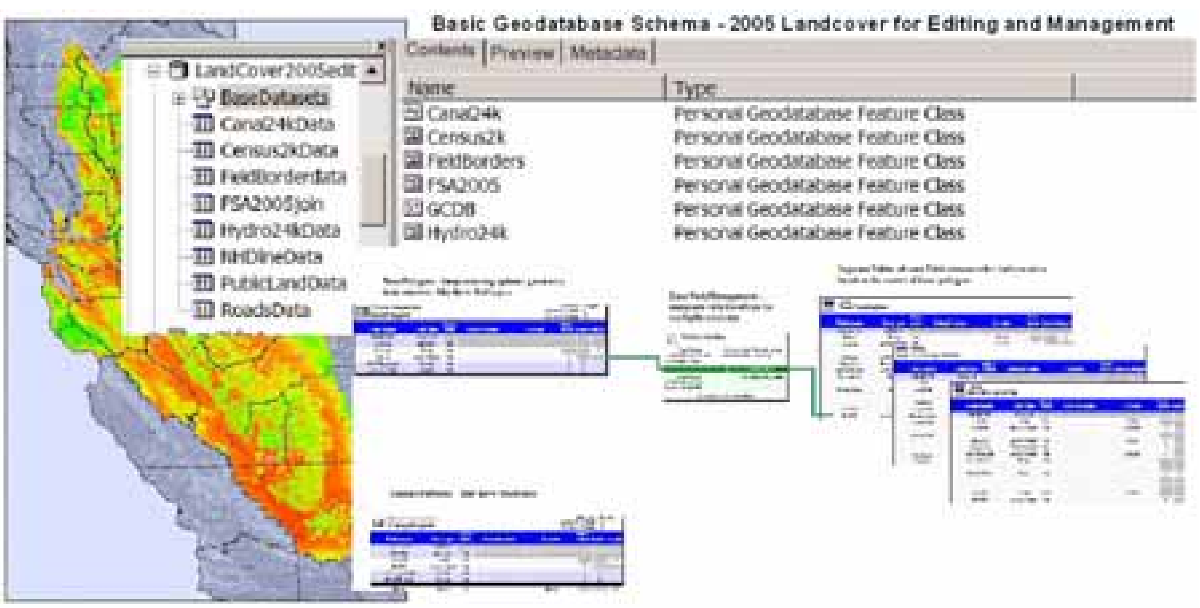
This management geodatabase is basically an editing geodatabase for capture of information from separate feature classes, updating feature classes, or adjusting the geometry of feature classes. It contains the following basic elements for the database schema.

- Simple Feature Classes
    - Polygons generated from multiresolution segmentation of Landsat 7 TM imagery for base polygons and change polygons
    - Polygons of 2005 crop mapping where available
    - Polygons of detailed hydrology where available
    - Polygons of wildlife areas and preserves where available
    - Polygons of Census tracts
    - Polygons generated from image segmentation and other source data such as urban rural residential areas

- Lines from road datasets
- Lines from hydrology including the canal system
- Points from GCDB for coordinate control and adjustments related to the Public Land Survey System

- Simple Table Datasets
  - 2005 Land cover classification for base polygons.
  - 2000 Land cover classification for base polygons.
  - 2005 Land use categorization as needed for rural residential, road networks, and hydrology networks.
  - Source attribute codes.
  - Tables generated by zonal statistics of polygon feature classes and raster datasets.
- Domains
  - Land cover classification for 2005
  - Land cover classifications for 2000
  - Land use categorizations for 2005
  - Data sources
  - Name of reviewer / editor

Relationship classes between the feature classes and table datasets are generally simple peer to peer relationships. Most will permit joins based on the source of the feature and the unique identifier for the feature. One exception to this is the crop data for 2005 from FSA. This is a one to many relationship where one field can have multiple crops with different planting dates. Topology rules are set within this schema to assist in merging and managing overlapping polygons generated from adjacent Landsat scenes. The basic schema for this managing and editing geodatabase is shown in Figure 8.



Feature class tables are generally kept very simple. Two attributes are common to all feature classes. One identifies the source of the feature. The other is a unique identifier for the feature. The related tables contain a subset of attributes from the source data that are useful for 2005 land cover mapping. This restricts the set of attributes loaded from the source data. Generally, only a subset of the source data is loaded into the feature classes.

This represents an implementation schema that will be reviewed and modified as the processing for 2005 continues. The review and modification of this model will follow the basic process identified in *Designing Geodatabases* by Arctur and Zeiler (2004), Associated with the datasets in the current personal geodatabase implementation are sets of raster datasets. With the release of ArcGIS 9.2, this model will migrate into a file based geodatabase that also contains the raster datasets. The model is currently being tested with the Beta version of ArcGIS 9.2. The use of geoprocessing tools and scripting reduces the frequency of locks preventing concurrent access to the database. This was also an issue in the development of the 2000 land cover dataset where the data was stored in ArcInfo coverage format. Access problems for review and editing should be simplified in the geodatabase structure where only subsets of a feature class or set of tables are needed at any one time.

Summary

A geodatabase is being used for management and editing of base polygons generated from multiresolution image segmentation and other source data to generate a 2005 land cover product. The data product will consist of polygon feature classes identifying land cover and change areas with associated tables. The geodatabase model with the use of geoprocessing tools assists in managing the datasets of feature classes, tables, and rasters used or produced in a complex series of processing steps.

This geodatabase ensures that all datasets have the same underlying spatial reference system. It also assists in any spatial adjustments that are required. The model consists of simple datasets of feature classes and related tables. Feature classes within the geodatabase contain a subset of attributes which uniquely identify the feature and data source. Associated tables in the geodatabase identify land cover information associated with the feature as well as key attributes from the orginal data sources. This assists in updates to the base polygons for the 2005 data product. Generally, it also permits simple table joins for one to one and many to one relationships. The model provides a framework for identifying topological relationships and rules for merging adjacent and overlapping feature classes. The geodatabase model provides a framework for the development of geoprocessing scripts to handle common processing tasks. The use of database domains assist in the use and application of common terminology for land cover and related land use categories. It provides a common framework for storage and maintenance of the datasets. With the release of ArcGIS 9.2, this model will be a file geodatabase which will include associated raster datasets. As an implementation model, it is expected to change as the process evolves.

References

Arctur, David and Michael Zeiler, *Designing Geodatabases*, ESRI Press, Redlands, California, 2004.

Curlis, Chris, David T. Hansen, Barbara Simpson, Demetria Adams, *Change in the Central Valley, California, 1993 to 2000* **Twenty Third Annual ESRI International User Conference**, San Diego, CA, July, 2003.

Hansen, David T., Barbara Simpson, Chris Curlis, and Jeff Milliken *Legend Development for a Land Cover / Habitat Classification Project for the Central Valley of California*, **Twenty First Annual ESRI International User Conference**, San Diego, CA, July, 2001.

Hansen, David T., Chris Curlis, and Barbara Simpson, (2006), *Techniques for Discrimination between Agriculture and Similar Land Cover Types with Fuzzy Logic and Spectral Polygon Characteristics* **ASPRS 2006 Annual Conference** Reno Nevada, May 3, 2006.

Simpson, Barbara, Chris Curlis, and David Hansen, (2006), *Identifying Rural Residential Areas within Land Cover Monitoring for California's Central Valley* **ASPRS 2006 Annual Conference** Reno Nevada, May 3, 2006.

University of Arizona, *MODIS Product MOD13A3 User Guide*, http://tbrs.arizona.edy/project/MODIS/UserGuide_doc.php, 2005.

Appendix

Datasets used in the development of land cover data for 2005. Most of this data cover the full study area, although they may not be complete for the full area. Some datasets also represent different time periods. Datasets include:

Land cover data for 2000

The land cover and habitat mapping for 2000 developed by Reclamation.

MODIS NDVI

MODIS NDVI is a data product available from NASA at two week intervals. The MODIS products used in development of the 2005 land cover product are 250 meter raster datasets from February 2005 through November 2005.

Census 2000 Data

U.S. Census tract boundaries and related data on urban clusters.

FSA Crop Data

USDA Farm Services Agency (FSA) data of field borders and reported crop and planting date for 2005 and other available detailed crop mapping.

NAIP Imagery

USDA National Agriculture Imagery Program (NAIP) 1 meter color orthoimagery for 2005 of California.

Hydrography

USGS National Hydrography Dataset (NHD) and related detailed hydrology data based on 7.5 minute topographic maps. This data is time dependent related to the last update to the 7.5 minute base.

Roads

Road network based on the U.S. Census TIGER line files and related detailed road network data.

Public Land Information

Identification of public lands in the area and the underlying coordinate base and control from the Public Land Survey System (PLSS) and Geographic Coordinate Database (GCDB) of BLM. These include natural areas such as wildlife areas and preserves.

Authors:

David T. Hansen
GIS Specialist / Soil Scientist
MPGIS
Phone: (916) 978-5268
FAX: (916) 978-5290
Email: dhansen@mp.usbr.gov

Chris Curlis
Remote Sensing Specialist
Division of Environmental Affairs
Phone: (916) 978-5030
Email: ccurlis@mp.usbr.gov

Barbara D. Simpson
GIS Specialist
Division of Environmental Affairs
Phone: (916) 978-5031
Email: bsimpson@mp.usbr.gov

Chuck Johnson, GIS Program Manager
U.S. Bureau of Reclamation
Mid-Pacific Region
2800 Cottage Way
Sacramento, CA. USA 95825-1898
Phone: (916) 978-5266
FAX: (916) 978-5290
Email: cbjohnson@mp.usbr.gov